Predictive Analytics Tools and Techniques

Mr. Chandrashekar G¹, Dr. N V R Naidu² and Dr. G S Prakash³

¹Student, M. Tech, Department of Industrial Engineering & Management, M S Ramaiah Institute of Technology, Bangalore, Karnataka
²Professor & Head of Department, Department of Mechanical Engineering, M S Ramaiah Institute of Technology, Bangalore, Karnataka
³Professor & Head of Department, Department of Industrial Engineering & Management, M S Ramaiah Institute of Technology, Bangalore, Karnataka

E-mail: ¹026chandrashekar@gmail.com, ²nvrnaidu@gmail.com, ³prakash5636@yahoo.com

Abstract

Predictive Analytics is the decision science that eliminates guesswork out of the decision making process and applies proven scientific guidelines to find right solutions. The central element of Predictive Analytics is the predictor, a variable that can be measured and used to predict future behaviour. The Predictive Analytics generally incorporates the steps of Project definition, Data collection and Data understanding, Data preparation, model building, Deployment and Model Management. Predictive Analytics is the form of data mining concerned with the prediction of future probabilities and trends.

In manufacturing sector, Predictive Analytics is an essential strategy to improve customer satisfaction by minimizing downtime while reducing service and repair cost. The data itself is getting smarter, instantly knowing which users it needs to reach and what actions need to be taken with the aid of Predictive Analytics. The technique through the collection, ingestion and persistence helps to uncover and pinpoint the failure patterns and build causal relationship over a large population of equipments leading to unit level failure prediction.

In the study, the focus is on the analytical tools that enable a greater transparency to analyse past trends and to predict the probable future rejection rate of the end product and how it impacts the rolling throughput yield metrics in manufacturing plant. The study provides the information about which statistical techniques can be applied for Predictive Analytics. Also, the illustration of results obtained from soft
computing tool that aid in performing Predictive Analytics. The application of Predictive Analytics will result in the increase of Rolled Throughput Yield (RTY) by 25% in the plant.

**Keywords:** Predictive Analytics; Regression; Soft computing tools; Rolling Throughput Yield (RTY).

### 1. Introduction

The predictive Analytics is one way of doing right thing first time. Debahuti Mishra et.al [1] defined the Predictive Analytics as the branch of data mining concerned with the prediction of future probabilities and trends. It focuses on predicting the future behaviour and occurrences of events based on the past trend. It serves as a tool in determining the critical process in a series of operations where the actual light has to be thrown in order to bring it to normal condition. In manufacturing sector, Predictive Analytics provides ideas on the occurrences of future downtimes and rejections thereby helps in taking preventive actions before abnormalities occur. As the number of rejection at each process in production affects Rolled Throughput Yield (RTY), there is a need to analyze and identify the critical process with the aid of statistical techniques and soft computing tools to take necessary actions. Craig Gygi et.al [2] suggested that RTY can serve as baseline scores and final scores for six sigma projects. Rolled throughput yield (RTY) is the probability that a single unit can pass through a series of process steps free of defects.

### 2. Material and Methodology

Predictive analytics is performed with the aid of statistical techniques. Few commonly used techniques are Regression Analysis, Time series Analysis, Factor Analysis, Artificial Neural Network, Decision trees, Naive Bayes etc. In the study, we use Regression tool for predicting rejection from each process.

Linear regression is the form of regression analysis that can be applied in various practical applications including manufacturing, business analytics and marketing. In general, it is used to estimate the unknown effect of changing variable over another. In specific, it models Y as a linear function of X. It is expressed as a straight line equation:

\[ Y = a + bX \]  \hspace{1cm} (1)

Where \( a \) and \( b \) are the regression coefficients.

‘a’ is the Y intercept and ‘b’ is the slope of the fitted line.

Linear regression model identifies the relationship between single predictor variable and the corresponding response variable when all other predictor variables in the model are held fixed. According to Doreswamy et.al [4], Regularisation of Linear Regression Techniques can make better the predictive error of the model by lowering the variability in the measures of regression coefficient by shrinking the estimates towards zero.
3. Overview of Turbine wheel and Shaft production process

Producing a shaft and wheel to meet specific demands is a technique that has been refined over many years. Fig1. Describes how the shaft and turbo wheel are produced and assembled. The focus of study is on Shaft production which begins from forged steel. The forged steel is subjected to friction welding, the process in which the friction between rotating and a stationary component causes the two metals to become red hot and forged by applying pressure. Welding is followed by turning of the shaft diameter in a lathe prior to precision grinding. Before grinding, the bearing diameter of the shaft is induction hardened. Final grinding of the shaft is done on CNC machines. Quality is assured through multiple function electronic gauges. The turbine wheel profile must be machined on the grounded shaft. Next seal the ring grooves on the hub of the shaft. Threads are rolled on the impeller end of the shaft. The final manufacturing operation is to assemble and balance the turbine wheel and shaft that is capable of running at an operating speed without vibration. According to link [5], almost all the dimensions on the shaft are critical to turbocharger performance and durability.

4. Cause and Effect Diagram to Identify Critical Process

The cause and effect diagram shown in Fig 2, it can be observed that the process 1, process 2, process 3 and process 4 are having effect on the final output, Rolled Throughput Yield (RTY). The contribution of each process to RTY can be identified by predictive analytics and it provides the information on which process is having
more effect on RTY. According to Anuradha R Chetiya et.al, [3], the process capability studies of Rolled Throughput Yield (RTY) can help determine if new equipment is capable of meeting the requirements and can also compare the capabilities of alternative equipments or machines. Also helps to focus on that critical process instead of analysing each and every process which saves time and capital.

Fig. 2. Cause and effect diagram for Shaft Production Process

5. Steps For Performing Predictive Analytics
The predictive analytics in the study is carried out by performing the below mentioned stages. Each and every step has its own significance and they impact on the final outcome.

1. Project definition: Defining the project objectives and desired outcome. The objective was to identify the critical process and predict future rejections to take necessary action and to maximise Rolled throughput yield (RTY).
2. Data Collection and Understanding: Related data about rejections and yield are collected for 3 months with two shifts a day. The data is then carefully examined to understand the behaviour of the equipments.
3. Data preparation: The next step is to prepare data for analysis by excluding unnecessary and unessential ones by making certain assumptions.
4. Model Building: The regression model is then built with Rolled Throughput Yield (RTY) as response variable and rejection at each process (process 1, 2, 3 and 4) are taken as predictor or explanatory variable. The linear regression model with high $R^2$ and high adjusted $R^2$ confirms the good fit. The model is built with the aid of Minitab software.
5. Deployment: Deployment stage involves the application of the fitted model into decision making process in industrial applications and to take the necessary actions for the critical process that are identified from the model.

For the study, the statistical computation is performed through Minitab statistical software. It makes it easier than ever to perform regression analysis and understanding results. According to Ginger Holmes Rowell et.al [6], the statistical
analysis computer applications have the advantage of being accurate, reliable and generally faster than computing statistics and drawing graphs by hand. Minitab is relatively easy to use once a few fundamentals are known. Minitab through statistical modelling determines which variables are related to a response and by how much. It also used in calculating expected values and forecast the impact of future changes.

7. Results and Discussion
The Rolled throughput Yield (RTY) from the past 3 months data for shaft production process was found to be 44.325%. The Rolled Throughput Yield includes the consideration of rework too as shown in eq. (2 and 3). The yield at $n^{th}$ process, Yield ($n$) is given by,

$$Yield \ (n) = \frac{Output \ from \ nth \ process}{Output \ from \ (n-1)th \ Process+\text{rework} \ materials} \quad (2)$$

And Rolled Throughput Yield (RTY) is given by,

$$\text{(RTY)} = \left( \text{Yield of process}1 \ast \text{Yield of process} \ 2 \ast \ldots \ast \text{Yield of process} \ n \right) \text{Where, n is number processing stages} \quad (3)$$

From calculation, the Average Rolled Throughput Yield for 170 shifts of shaft production was found to be around 45%. The Regression model is fitted by taking the response variable as Rolled Throughput Yield (RTY) and the predictor variables as the number of rejections at process 1, process 2, process 3 and process 4. The Regression equation obtained is given by,

Since, $R^2$ and adjusted $R^2$ are considerably high i.e, 80.82% and 80.36% respectively, it is found that the regression model fitted is a good fit. Hence, this model can be accepted for prediction based on yield observation. $R^2$ is the coefficient of determination which measures how well the least square regression line fits the
sample data. The main purpose of $R^2$ is to either predict future occurrences or test of hypothesis.

Fig. 4 Snapshot of Anova from Minitab software

Fig. 5 Snapshot of Pearson Correlation from Minitab Software

Fig. 6 Scatter plot of Rolled Throughput Yield v/s rejections from process 1, 2, 3 and 4
The scatter plot describes the possible relationship between RTY and rejections from each process. It can be observed that there is a negative correlation between RTY and rejections from the above graphs. Since, there is a clustering of data around the fitted model, it is clear that there is a strong relationship among the variables in spite of considerable amount of unusual observations. Pearson correlation of RTY and rejection for process 3 is -0.013 which represents that there is a slight negative effect of process 3 and no need to focus more on process 3 factors as they are well within control.

By the above study, it is clear that the fitted model is a good fit. By regression equation, it is found that the process 1 which includes casting and welding appears to be most critical and the necessary actions are to be taken to control the process by carefully examining the critical factors in process 1. Process 2 which involve induction hardening and grinding also needs to be focussed and the necessary actions in order to control the critical factors are to be taken. There is no need to focus on process 3 of wheel profile machining and process 4 of Threading and Sealing as they are not affecting the Rolled Throughput Yield (RTY) considerably. Since, the model relating RTY and various processes are fitted, the rejections at each inspection stages can be obtained for the expected RTY of 70% which is 25% more than current value. This improvement can be obtained by focussing on key issues that can be identified through Design of Experiment analysis for process 1 and process 2 which are affecting on output.

8. Scope for Future
The study can be further enhanced in the following dimensions:

1. To perform Design of Experiments to find out the factors those are affecting the critical processes by factor analysis and Principal Component Analysis.
2. To implement the Industrial Engineering Techniques to bring the abnormal factors to normal conditions and to run the pilot study.
3. To calculate the Rolled throughput Yield of the pilot study and compare it with the current value.

9. Conclusion
The study pursued to investigate and predict the future rejections from a particular process if the expected Rolled Throughput Yield (RTY) is known. The predictive regression model helps in identifying the critical process and to take the necessary actions as required. The identification of critical process becomes easier and cost effective as only statistical techniques and tools are used. Instead of investing each and every process, which is time consuming and complex, Predictive Analytics helps in identifying and focussing on only critical process. The study becomes the predecessor for Factor Analysis if the critical process is identified. By predicting the future occurrences of rejections, it is helpful to take the necessary actions before abnormalities occur.
References


